Introduction

- Constructed-response assessments reveal student thinking and conceptual barriers.
- Automated analysis allows constructed-response items for JITT in large classes.
- Our approach to automated analysis is shown in Figure 1.

Sample assessment: Functional group question

Consider two small organic molecules in the cytoplasm of a cell, one with a hydroxyl group (-OH) and the other with an amino group (-NH2). Which of these small molecules (neither or both) is most likely to have an impact on the cytoplasmic pH?

A. Amino B. Hydroxyl C. Both D. Neither

Please explain your answer.

Word count applications fail to reveal complex concepts in responses

Discriminant analysis can create classification functions.

- Identified the most important 6 categories for predicting the human rating (see highlighted categories in Figure 3).
- Functions predict human score of student response with 77% accuracy (Table 2)
- Computer – Human Inter-rater Reliability = 0.835

Lexical analysis can categorize large number of student responses easily.

- Expert input required to customize libraries and develop categories.
- Categories contain multiple terms and rules.
- Responses can be included in multiple categories.
- Output includes a variety of visualizations of responses (e.g. Figure 3).

Acknowledgements

We thank Ryan Sweeder for help in scoring student responses. This work was supported by the Center for Research on College Science Teaching and Learning (CRCSTL) at Michigan State University and the National Science Foundation (NSF; awards 0736952 and 1022653). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

Table 1. Scoring rubric used to rate student explanations. Number of correct multiple choice responses at each scoring level are indicated, along with an example student response at each level.

Table 2. Classification percentages of cross-validated student responses for functional group classified at each level.

Table 3. Distribution of responses in each category. Categories identified as significant in the scoring prediction function are highlighted in green.